Accepted for publication in Archives & Manuscripts on 18 Nov 2017

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

Records storage in the cloud: are we modelling the cost?

Julie McLeod and Brianna Gormly

Julie McLeod, iSchool, Dept. Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, UK

Brianna Gormly, Archives and Special Collections, Martin Library of the Sciences, Franklin & Marshall College, Lancaster, USA

Abstract

Cloud services are increasingly seen as a flexible strategy for platform, infrastructure and software. Given the cloud's claimed economic benefits archives and records professionals are now using cloud services for the storage of digital records and data. However, in determining whether or not to use the cloud for records and/or data storage, what models are available to them for estimating the cost and the medium to long-term financial implications for their organisations? This article identifies models available for estimating cloud storage costs and presents the results of an international survey into their use in the decision-making process with a series of real-use case examples illustrating their value. The study highlights a series of important implications for archivists and records managers. These include the importance and challenges of using the models, their lack of widespread use, their adequacy, and the multiple players who should be involved in their application and development. Archivists and records managers need greater awareness and understanding of the models so they can play a central role in the cloud storage decision-making process and in the development of more effective costing models.

Introduction

Cloud services are increasingly seen as a flexible strategy for platform, infrastructure and software services. Indeed, International Data Corporation anticipates worldwide expenditure "on off-premises cloud IT infrastructure will experience a five-year compound annual growth rate … of 14.2%, reaching $48.1 billion in 2020", by which time the total spend on cloud IT infrastructure (both on and off-premises) will exceed the spend on non-cloud IT infrastructure.[1] As more data and records are generated, the economic sustainability of using cloud services for its storage is an issue that needs the attention of recordkeeping professionals. This article presents the findings of research that explored models for costing cloud storage services and their application (or not) in practice. It first provides background on the push toward cloud storage of records and a literature review discussing previous research into costing models in the context of storing data and information in the cloud. Responding to a gap in the literature, the remainder of the article discusses a global empirical study of the use of these models in practice that focused specifically on the storage of organisational information i.e. records. The findings are presented and their implications for Archives and Records Management (ARM) professionals are discussed. The article concludes by suggesting better models are needed and ARM professionals need to position themselves more effectively in the decision-making process.

Background

Cloud storage is important because it can improve access to and sharing of records and data, increase security, increase potential for long term preservation, take advantage of

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

economies of scale and (potentially) save money. Literature from service providers and consultancy companies has highlighted the economic benefits of using cloud services for the storage of digital information.[2] On what basis are these claims made and are cloud storage services economically sustainable? In a records management context, Stuart and Bromage noted that cost presents a significant pull toward cloud storage and there is now evidence that archivists and records managers are increasingly using the cloud for the storage of digital collections.[3] Digital records may be stored in the cloud as part of an intentional archives storage program, or business systems may store current records in the cloud by default. Regardless of duration, cloud storage represents a new, ongoing cost which ARM professionals must fully understand and carefully consider.

Decisions about in-house versus cloud storage are complex; there are big implications if the wrong decision is made. Models can be helpful in arriving at a decision, systematising a highly complex decision-making process. While organisations may explore a range of methods to assess cost, using a generally accepted model means decision-makers can have more confidence in their decision. What models are available to them for estimating the cost and medium to longer term financial implications for their organisations of moving the storage of some or all of their digital collections to the cloud and how are these models being used? These are important and urgent questions in the context of trust in the economic viability and sustainability of using cloud storage for digital information.

For records and archives collections, these questions are particularly significant since they are inherently unique, provide evidence of business activities, and are also information assets with special characteristics.[4] Given their complexity, digital records can be challenging to store, especially through time, being dependent on (changing) software/hardware and requiring sufficient metadata for management and preservation. They are, arguably, the most complex form of recorded information organisations have to manage. It is, therefore, important to investigate decision-making models for cloud storage and, where there are gaps, for the profession(s) to develop such models. Developing models to address cloud storage decisions for records could serve as the basis for the development of models that would support cloud storage decisions regardless of the form or type of recorded information or the nature of the organisation.

Literature Review

A critical review of a purposive selection of literature across multiple disciplines, covering information management, records management, archival science, computer science and IT, and business, as well as the websites of relevant organisations, was conducted. Details of the search strategy, results and selection of sources are given in McLeod and Gormly.[5] The review identified a body of work on modelling the cost of data and/or information storage in the cloud, which was divided into three tiers based on its relevance to the research objectives. Only the most relevant sources are discussed here.

Relevant previous research on modelling cloud storage costs

Extending the early work of Walker,[6] Walker, Brisken and Romney applied a pre-existing buy-or-lease business decision model to the use of the cloud for storage infrastructure in three hypothetical scenarios.[7] They sought to ‘stimulate discussion, debate, and future work in the quantitative modelling of the cloud computing industry’ and ‘to assist consumers,

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

researchers, and policy makers in estimating the benefit of leasing from storage clouds'.[8] Subsequent work addressed the model's perceived weaknesses resulting in probabilistic models that assess the risk of making the decision.[9] This work is cited by Mazhelis, Fazekas and Tyrväinen and by Laatikainen, Mazhelis and Tyrväinen,[10] although their model has a different basis. Other relevant scholarly work includes Khajeh-Hosseini et al., Wang et al., and Dutta and Hasan, with Reichman presenting the work of consultancy company Forrester Research Inc.[11] All of these authors are situated in the computer science or information systems disciplines, with the exception of Mastroeni in economics.

Relevant research is also found in the library and archives disciplines, focusing on digital preservation. Particularly important is the work of Rosenthal and Vargas, D C Rosenthal et al., D S H Rosenthal et al. and the 4C (Collaboration to Clarify the Costs of Curation) project's 'Evaluation of Cost Models and Needs and Gap Analysis'.[12] The latter includes both DC and DSH Rosenthal et al.'s economic model for long-term preservation storage.

Four different financial or management accounting theories, with some 'variations on a theme', underpin the models presented in this work, namely: 1) Discounted Cash Flow including Net Present Value, Differential Net Present Value (DNPV) and Internal Rate of Return (IRR); 2) Monte Carlo models and Kryder's Law; 3) Full Cost Accounting including Total Cost of Ownership; and 4) acquisition intervals for additional storage. Table 1 summarises the key characteristics of these theories, referencing the authors who have published about their use in the cloud storage context. Further details can be found in McLeod and Gormly.[13]

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

Model / Financial Theory	Characteristics	Authors (chrono)
Discounted Cash Flow (DCF) including: - Net Present Value (NPV) - Differential Net Present Value (DNPV) - Internal Rate of Return (IRR)	DCF: based on the principle of the value of money (spent or invested) over time; i.e. a unit of money today having a different value in the future, taking account of inflation, interest rate (the discount rate) and returns. Standard economic techniques but sometimes criticised because they assume the interest rate is constant rather than variable over time. NPV: sum of the present values of all the cash flows relating to a project, i.e. cash inflows (earned) and cash outflows (spent). A positive NPV indicates a profit, a negative NPV a loss. In a buy-or-lease scenario if NPV(buy) is greater than NPV(lease) then the decision should be to buy. Considers factors such as capital costs (e.g. purchase, interest rate), operating costs (e.g. energy, personnel), and other relevant costs for the context (e.g. disc price trends, disc replacement rates and hardware salvage value). DNPV: considers the difference between the two NPVs rather than their absolutes; easier to calculate. IRR: the interest rate required for the NPV to be zero.	Walker, Brisken, & Romney (2010); Mastroeni & Naldi (2011a); Mastroeni & Naldi (2011b); Wang et al. (2012); Naldi & Mastroeni (2013); Naldi & Mastroeni (2016) See also: Khajeh-Hosseini et al. (2012); DSH Rosenthal & Vargas (2012); DC Rosenthal et al. (2012)
Monte Carlo / stochastic models and Kryder's Law	Monte Carlo models: based on Monte Carlo or other stochastic methods i.e. algorithms that use statistical probability to handle uncertainty (e.g. future storage costs or volumes) Kryder's Law: states that storage density of discs doubles every two years; widely translated into the exponential decrease in digital storage cost.	DSH Rosenthal et al. (2012); DSH Rosenthal & Vargas (2012); Naldi (2014);
Full Cost Accounting (FCA) including Total Cost of Ownership (TCO)	FCA: recognises a wider range of costs than standard financial cash flow methods, e.g. economic, social and environmental costs. TCO: sum of all expenditures of a project or system (e.g. power, personnel, hardware), accounting for direct and indirect costs, including overheads, but not for the time value of money. Can be used in FCA.	Reichman (2011); Dutta and Hasan (2013)
Acquisition interval	Length of time between intervals at which an organization evaluates its storage needs (including predicting demand for storage) and acquires additional in-house storage.	Mazhelis, Fazekas, & Tyrväinen (2012); Laatikainen, Mazhelis & Tyrväinen (2014)

Table 1. Summary characteristics of costing models used in a cloud storage context featuring in the literature

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

Models for digital storage costs

The Discounted Cash Flow models are potentially less useful for modelling digital storage costs over the longer term, according to Walker, Brisken and Romney, since they rely on past data.[14] However, their more sophisticated versions (DNPV, IRR), which take account of unknown or random changes (for example in leasing price and disc failure) and incorporate risk measures, can be useful. The probabilistic nature of Monte Carlo models are also helpful for longer term cost modelling of data or records storage as they too account for unknown changes (e.g. interest rates). An important factor here is Kryder's Law,[15] which states that storage density of discs doubles every two years, though it is widely translated into the exponential decrease in digital storage cost. Work on projecting cloud storage costs and the impact of Kryder's Law and other factors on the costs, by Rosenthal et al. and Rosenthal and Vargas revealed that cloud storage pricing has not decreased according to this law.[16] In fact, pricing of some of the major cloud storage providers had dropped at most by 3% per year or remained the same in comparison with a 30% per year drop in disc prices over a 30 year period.

Total Cost of Ownership (TCO), Reichman argues, is difficult to use accurately in practice and suggests 'a more pragmatic approach is to compare only the costs that change between the two scenarios, known as relative cost of operations'.[17] Changing factors include service life of storage; storage acquisition cost; redundancy copies; storage utilisation; personnel; infrastructure cost (facilities and energy); maintenance and data migration. The final set theories accounts for acquisition intervals for additional storage, considered by some to be critical in analysing storage cost.[18]

Modelling digital storage costs in practice

Only three 'real' case examples of modelling digital storage costs in practice were found in the literature, all of which are in university contexts. In UK and US settings respectively, Khajeh-Hosseini et al. and Dutta and Hasan found little difference between the cost of in-house servers and cloud storage but highlighted the need to consider factors beyond financial considerations, such as organisational change, pricing, scale of operation and data redundancy.[19] In contrast, Mazhelis, Fazekas and Tyrväinen and Laatikainen, Mazhelis and Tyrväinen suggest that, in a typical case of exponentially growing storage demand, public cloud storage is more cost effective when the acquisition interval assessment of in-house storage capacity is longer.[20] However, this interval is affected by other factors (for example necessary storage redundancy, and the cost of transferring data to and from the cloud).

Five hypothetical scenarios provide comparisons based on the relative size of the organisation irrespective of sector.[21] These mostly indicate that cloud storage is more cost-effective in the shorter term (less than ten years) and purchase is a better long term (ten or more years) investment, particularly for larger companies. Three further scenarios provide comparisons using service providers' pricing structures.[22] Interestingly, Rosenthal and Vargas concluded that, because cloud storage pricing has not decreased according to Kryder's Law, the in-house option is cheaper for long-term storage.[23]

In summary, since Johnson and Lewellen's seminal paper modelling the buy or lease decision,[24] a range of increasingly sophisticated models have been developed to understand and

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

compare the cost of storing data and information in the cloud versus in-house. However, the models have limitations and the authors discussed above note caveats regarding the conclusions that can be drawn from their work. With one exception, in which Dutta and Hasan based in computing and information sciences cite DSH Rosenthal et al. from library/archival science, [25] there is little citation between the complementary work that has been undertaken and published in these two disciplines. This suggests the work has been undertaken in parallel 'silos'. If this separation plays out in practice then there is a danger that information professionals may not be cognisant of the economics of cloud storage, and hence not consider this with their computer science or information technology colleagues, and vice versa. The low number of published case examples is also notable. This raises significant concerns regarding whether or not information professionals currently have adequate information to make an informed decision about using the cloud to store some or all of an organisation's data, information and/or records.

The Research

For recordkeeping professionals, the issue of using cloud services to store records and archive collections is particularly important. If they are to be used, their viability, sustainability and trustworthiness are paramount. However, no published case examples exploring these issues in a recordkeeping specific context were identified. An empirical study was therefore conducted to address this gap by posing the following research questions:

- What economic/costing models are used in practice and why/why not?
- How are economic/costing models used in practice?
- What other models/approaches are used to underpin the decision-making process about the use (or not) of the cloud for Storage as a Service (StaaS)? Why and how?
- How do the models/approaches used in practice compare with economic/costing models from the literature in terms of the factors considered?
- Who is involved in the decision-making process for using the cloud for StaaS and why?
- What are the issues of trust in using the cloud for StaaS and what factors contribute to trust in the decision-making process?[26]

The study was conducted primarily via a survey, using the online tool FluidSurveys,[27] towards the end of 2015 and again during February 2016 to improve the response rate. To target a global audience of recordkeeping professionals, it was disseminated through a purposively selected global set of archives, records and information management listservs, and forwarded by international research colleagues to contacts and relevant stakeholders in their organisations.

It was split into two sections, the first pertaining to the respondents' organisations and the second dealing with respondents' experiences in their specific roles within the organisation, endeavouring to distinguish any differences in perspective. It covered the use (or not) of third party cloud service providers for records storage and the use of costing/economic models in the decision-making process. All questions offered multiple choice responses.

This article presents an analysis of our findings on the use of cost models and their application in the decision-making process, and aims to explore our research questions related to cost modelling in practice. Full details of the survey data are available in McLeod and Gormly, including questions related to the wider socio-political, technological and organisational issues.[28]

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

That portion of the survey aimed to answer the final research question regarding issues of trust in cloud storage, a full discussion of which is beyond the scope of this article and addressed elsewhere.[29]

Respondent demographics

Sixty-one completed survey responses and 115 incomplete responses were received. Only the complete responses were analysed. Since the survey was disseminated to channels targeted at recordkeeping professionals, it is unsurprising that over half of respondents (54%) were recordkeeping practitioners. Of those who chose 'other,' half specified a combination of recordkeeping and other areas, including administration, IT, risk, Freedom of Information, and e-government. The remainder were in faculty, research, and executive roles. Respondents came from organisations based in 17 different countries. However, the majority were in English-speaking countries, 66% from Australia, Canada, the United Kingdom, and the United States. The largest group of respondents identified their organisation as a governmental one (33%), followed by education (23%).

The case examples

The survey was also used to identify organisations using costing models, or related approaches, that were prepared to share them publicly as case examples. Sixteen respondents offered to be interviewed and provide a case example, nine of whose organisations were using the cloud and/or a costing/economic model. A global sample of 5 were interviewed; they came from 4 countries in 3 continents, in different public, private and third sectors, and ranged in size from 250 to 3000 staff. They all used the cloud for records storage but only three had used a costing model in reaching the decision.

Findings - Survey

Use of a cloud service provider for records storage

Cloud services were defined here as including public, community or hybrid clouds as well as private clouds managed by a third party. Roughly half of the respondents (52%) said their organisation used such a provider, 43% did not and the rest did not know. Respondents in organisations that did or did not use a cloud service for records storage (58) were then asked why they made this decision, the reasons chosen from a list of options provided. The most popular reasons in organisations using cloud storage were cost savings in hardware and software (~78%) and in human resources (~44%) (Figure 1). In organisations that had chosen not to use cloud storage, lack of trust in cloud service providers was the most popular reason (50% respondents), followed by legal or regulatory requirements (38%) (Figure 2). This data supports the observations motivating the research that cloud storage is often promoted and chosen for financial reasons. However, it also indicates that trust and legal or regulatory context take primacy over cost savings for organisations not using cloud services.

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

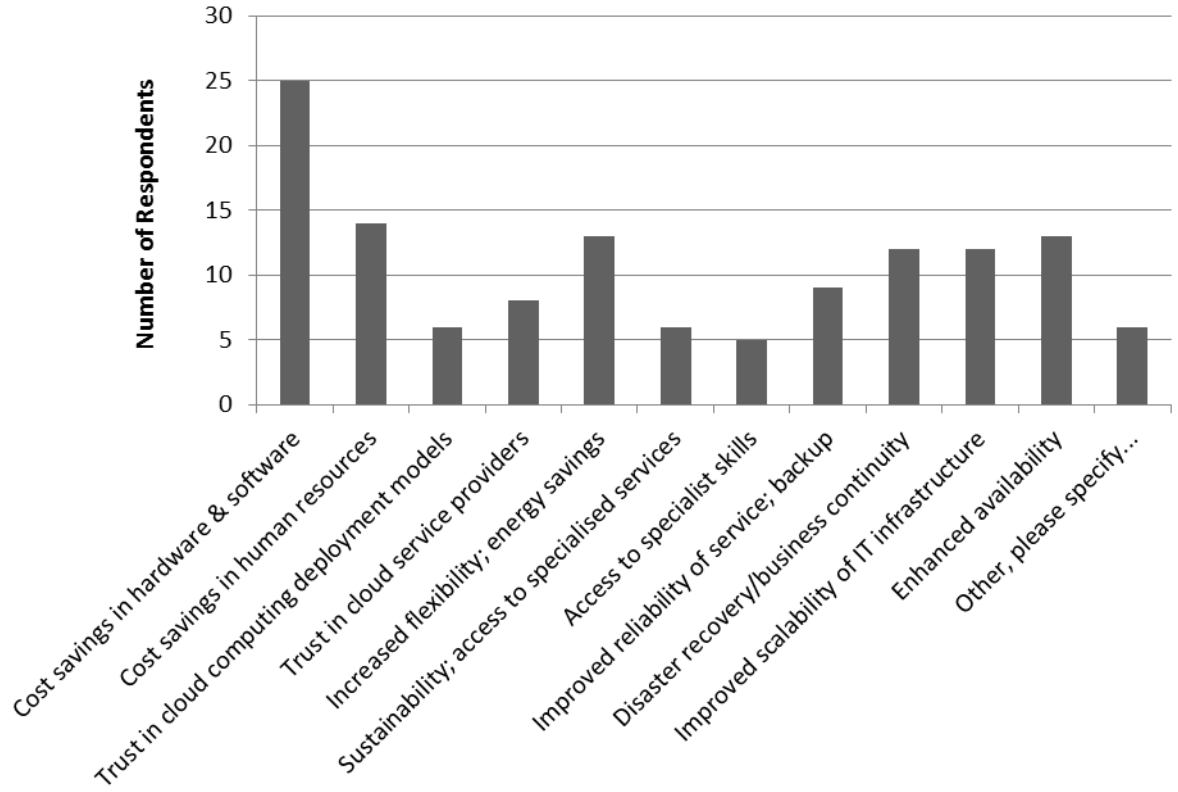

Figure 1. Reasons why organisations use a third party cloud service provider for records storage (No. of respondents = 32)

In addition, these 58 respondents were asked about the factors their organisation considered in the cloud usage decision-making process. Respondents could choose as many factors as were applicable, again from a list of options provided. The most frequently chosen were operating costs (41 respondents), technological suitability (37 respondents), risks (31 respondents), and capital cost (29 respondents). Further, they were asked specifically about the importance of cost as a factor in the organisation's decision-making process. In total 86% rated cost as either very important (38%), essential (28%) or critical (20%) for their organisation in the decision-making process, demonstrating that cost was clearly an important factor. Again, this data supports the observations that originally motivated the study regarding how the assessment of cloud storage appeared to be cost-driven.

For respondents in organisations using cloud storage for records, the survey also endeavoured to understand how the service was being used. Respondents were asked about the use of the cloud for short term (1 to 9 years) and/or long term (10 or more years) storage. A majority of respondents (~59%; 19 respondents) said that they used cloud storage for short-term storage of some records with 44% (14 respondents) using it for long-term storage of some records. Only 4 respondents used the cloud for short-term storage of all records, and no respondents used the cloud for long-term storage of all records. The remaining respondents chose either 'other' or 'don't know'. While this data is interesting in how it illustrates current use of cloud services, the small amount of data limits any meaningful analysis of correlations with cost concerns or cost modelling (discussed below).

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

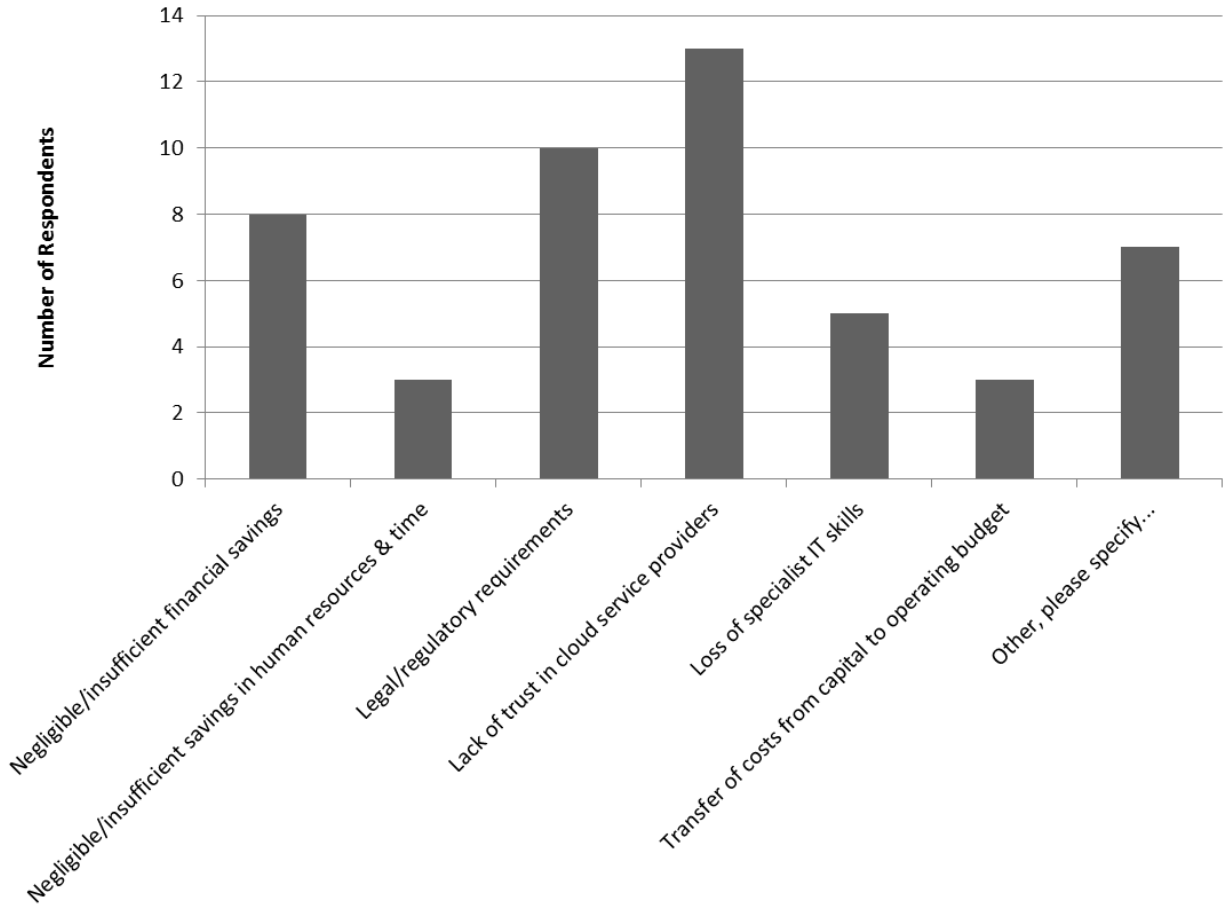

Figure 2. Reasons why organisations do not use a third party cloud service provider for records storage (No. of respondents = 26)

The use of costing/economic models in the decision-making process

About 25% (15) of the 58 respondents in organisations that did or did not use a cloud service provider for records storage indicated that their organisation had used a costing model in the decision-making process; of these 10 were using cloud storage. Regarding the specific model(s) used the most popular ones were relative cost of operations (10) and non-financial factors (9) (Figure 3). Few had used the more sophisticated models identified in the literature review, raising questions about the adequacy of their cost modelling. However, this potential issue may be countered by the respondents' use of multiple models. Of the 15 organisations one third used two models, one third used three or four models and the remaining third used only one model.

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

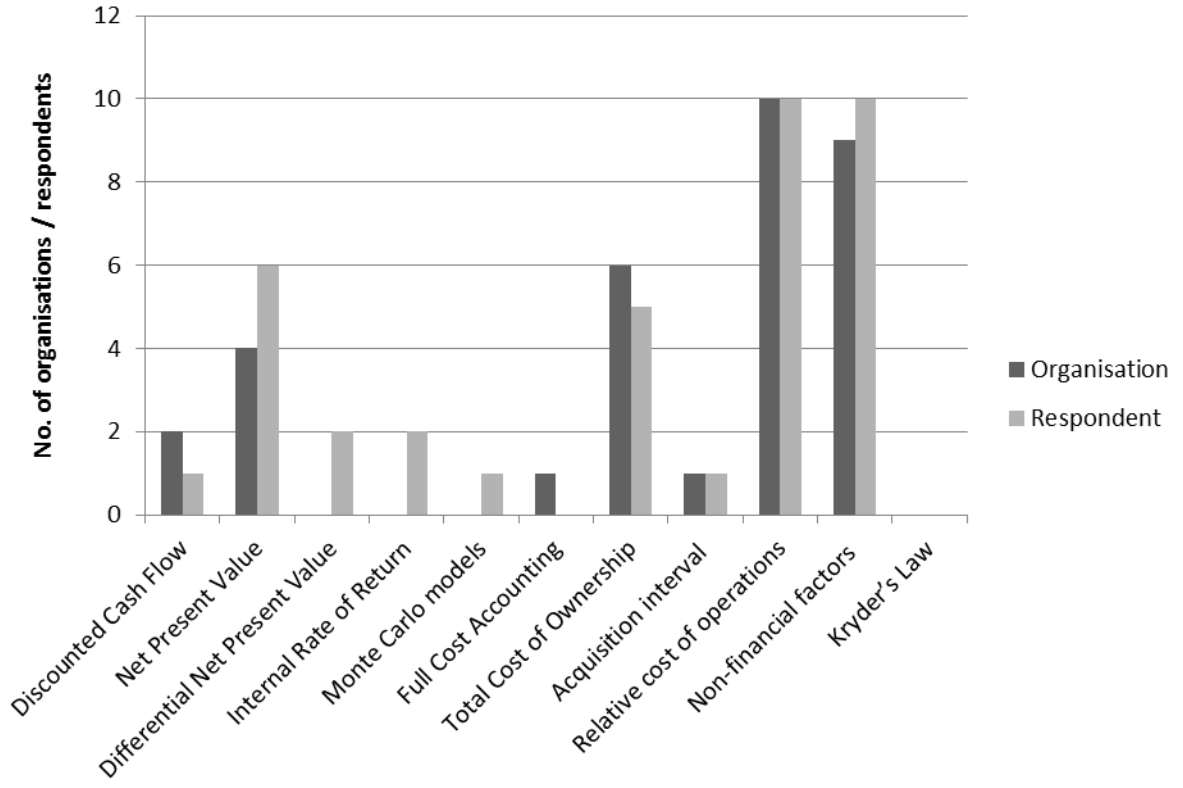

Figure 3. Costing/economic models used by organisations and respondents in deciding whether or not to adopt an external third party cloud service(s) for records storage (No. of organisations = 15; No. of respondents=11)

Respondents who indicated they were either largely or partly involved in the decision-making process (47) were also asked if they themselves had used such a model. Again, only a small number of respondents (11, ~19%) said that they had used one. The models most used were the same as at the organisational level (Figure 3). Whilst this reflects a similar tendency not to use the more sophisticated models, some of the survey respondents had used the latter themselves, specifically Differential Net Present Value, Internal Rate of Return and Monte Carlo models. Unlike their organisations, no respondent had used Full Cost Accounting and Kryder's Law had not been used by any respondent or organisation. No models beyond those included in the list of options provided were used.

The 10 respondents whose organisations used both a costing model and cloud storage were asked how both they and their organisation used the model(s) in the context of using a cloud storage service(s). Models were generally used for the same purposes at both the organisational level and in respondent's specific role,[30] the most popular being to estimate costs as part of the adoption decision-making process. However, this small amount of data does not offer insight into the rationale behind their use of particular models.

Roles and responsibilities

The survey also sought to gain an understanding of the various parties involved and their level of involvement in the organisations' decisions regarding cloud storage for records. Forty-

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

three respondents (~71%) selected the IT department, followed by archives/records management (38, ~62%) and senior management (30, ~49%). Other parties included administration, finance, information governance and/or security and legal. When asked which department played the lead role, IT was the most frequent (22) followed by a combination of more than one department (11), implying there was no single lead. The latter is positive in the sense that multiple stakeholders are involved in such decisions. In fact, three of these respondents indicated that both IT and the archives/records management departments were involved and another two specified a combination of information governance and IT. However, the fact that only four respondents (~7%) said that recordkeeping professionals played the lead role is worrying given the focus was on the adoption (or otherwise) of cloud services specifically for records storage. These numbers suggest that, while ARM professionals are often involved in the decision-making process, it is the IT department that generally plays the lead role.

Findings - Case examples

Organisational contexts

Table 2 provides brief contextual details for the five case examples. They all use the cloud for storing records but are diverse in how they use it. One organisation uses the cloud specifically for its digital archives accessible only by recordkeeping staff. In three cases, the cloud is used for business systems and, therefore, records are being stored. One case uses it for a specific business function (teaching) but not for its own organisational records. Together they illustrate the contrasting uses of cloud storage – on the one hand use is explicitly for cost efficient storage of archival records as part of a digital repository initiative and on the other (non-archival) records are being stored 'by default', as part of the use of cloud infrastructure and software services for business systems.

Case 1: Large UK public sector body
Developed its first digital repository in 2013, making use of cloud services for storing low usage archival records that have no security classification. Includes records of some core business functions but primarily archived websites and digitised archival records (public access is to other copies). Present volume ~17 Terabytes. Archives Department's decision to use cloud services motivated by the organisation's adoption of a 'cloud first' IT strategy in alignment with UK Government's 'Cloud First' policy, which potentially meant there would be no or few in-house data centre services in the future. An in-house digital repository storage centre would then bear a disproportionate level of inherent overheads, substantially increasing costs. Other reasons were: trust in cloud computing deployment models and cloud service providers; increased flexibility; enhanced availability; improved reliability of service; improved scalability of IT infrastructure; backup, disaster recovery/business continuity.
Case 2: Large 100-year-old Canadian technical college, satellite campuses around the world
Has a records management team and an archives unit. No formal records management program until 2015. Many hard copy records stored with a commercial service provider, large percentage of born-digital records being created and stored in digital form only. Organisation looked at the cloud for digital records storage and cost savings in human resources. Began using Apple's iCloud to store some records in 2015 through a time-limited introductory offer. Though not a solution to the lack of a fully developed records

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

management program, in conjunction with a new retention schedule and development of records classification, the cloud offered an alternative storage solution for inactive digital records. However, senior management put its adoption on hold.
Case 3: Large New Zealand state owned enterprise, many contractors Has a well-established records management service and is one of Microsoft's early adopters worldwide. Decided to move to an evergreen platform when Microsoft offered a big discount to move to its new cloud platform and services suite. Microsoft was looking to trial its new platform. Organisation was in a good position in terms of IT lifecycle management; offer showed a substantial monetary advantage against its 3-5 year budget, though the organisation recognised there would be risks. Cost was not the only driver for moving to the cloud; increased flexibility, access to specialised services, evergreen technical support, avoiding shadow IT (i.e. individual staff or business units 'doing their own thing') and gaining centralised control, ability to work collaboratively with third parties, better remote/home working support, 24/7 access and use of portable devices (part of business transformation) were other drivers. A strategic decision for organisational benefit but providing an excellent opportunity to move its records management to the cutting edge (e.g. implementing ontology driven records management with front end auto-classification).
Case 4: Large Spanish City Council responsible for governing the city, providing public services administration and fostering socio-economic development of the area Well-established ARM departments with records management processes/requirements well integrated into management and business systems. Has used cloud type platforms to provide and manage public services and projects for a long time. Now uses the cloud for Software as a Service (e.g. to maintain public street lighting, to manage incidents in collaboration with the Police Service). Using these cloud services means records are created, used and therefore implicitly stored within those systems. Present volume ~ 2 Terabytes. Started to use a cloud service to manage its own records in 2005. In principle does not use the cloud for storing its 60 Terabytes of digital archival records, these are managed in its own system.
Case 5: Medium Canadian religious organisation formed by an amalgamation of four related organisations, with staff spread across four different geographical locations Organisation does not have a records management program but has used public cloud services to store some records since 2014. Main driver for using the cloud was to solve problems of file sharing between different locations, provide access to files and email for leadership members and staff travelling on business, and connect everyone. Uses the cloud for generic office software and a professional association archives catalogue database. Administration, finance, personnel and facilities management records are stored in the cloud but firm decision not to store archival records in the cloud.

Table 2. Case example contexts

Roles and responsibilities

In all cases, the IT Department played the lead role in the decision-making process for using the cloud for records storage. In two cases this role was shared with the Archives/Records Management Department; however, in two other cases, recordkeeping professionals were not involved or consulted. In all cases, other players were involved, ranging from finance to senior management (who in some cases approved the business case or made the final decision), to

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

departments with specific needs or even the entire organisation. The involvement and roles of these players were based on a number of factors. In some cases, arrangements were simply part of normal organisational processes (e.g. cost modelling for IT services being led by IT). In others, roles were dictated by which parties were most directly involved in or most responsible for providing support. Organisational culture, or dynamics between the individuals involved, strongly influenced the situation in some cases. Post-adoption cost-monitoring responsibilities varied.

The use of economic models in the decision-making process

In the three organisations using economic models, the most popular models were Relative Cost of operations (3), non-financial factors (3) and Total Cost of Ownership (2), reflecting the trend in the survey responses. One organisation also used Discounted Cash Flow and Net Present Value. All but one used more than one model, mirroring the survey results. Overall, the reasons for using these models were either that it was a required part of the business case (*Case 1*), or that it met the particular decision-making concerns (*Case 2*), or it was standard practice (*Case 3*).

A focus of the interviews was identifying how these organisations used costing models in practice. In *Case 1*, the ICT Department used the Relative Cost of Operations model to compare the cost of in-house versus cloud storage and to consider how in-house storage services might develop and affect the cost of that option. The Archives Department also considered non-financial factors, in particular the risks and responsibilities related to fulfilling archival requirements, to ensure the organisation understood the decision was not only about storage cost but also a particular type of storage with particular records requirements. Being their first digital repository, there was no existing collection to use as the basis for modelling costs. However, robust data was available for volume, rate of acquisition and usage of (the priority) records, since this was a back-office (master) repository used only by the Archives Department. Modelling over an eight-year period showed that the cloud was the most economic option by some distance, and the Archives Department had separately assured themselves that cloud services could meet the non-financial (archival) requirements. As yet there had been no formal review of the original modelling against current reality, but costs were in line with what was expected.

For *Case 2*, although cost was very important and probably the main factor in the decision-making process, the IT Department's main concern, as the lead decision-maker, was whether or not cloud services could scale and meet future needs. They compared a number of services in terms of cost but did not use a purely financial model; they also evaluated non-financial factors, particularly technology suitability (i.e. functionality/features, industry/collaborator trends), stakeholder impact, risks and responsibilities. The cloud adoption decision-making process was informal and based on the organisation's past experience.

In *Case 3* the IT Department, which includes the records management function, uses Relative Cost of Operations and Total Cost of Ownership to model the cost of software applications. However, Total Cost of Ownership for on-premise solutions often have not included the cost of the people component of a service, in particular hidden or 'unconscious' costs (for example someone's time helping to 'get something done'). The organisation bears these hidden operational costs. Using Total Cost of Ownership in the cloud environment was proving to be challenging. Non-financial factors were also being considered.

Accepted for publication in Archives & Manuscripts on 18 Nov 2017
Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

Discussion

Use of economic models in practice

Overall, the clearest implication of the survey data is a lack of widespread use of costing/economic models, either by respondents (~19%) or their organisations (~25%), in deciding whether or not to use a third party cloud service provider for records storage and/or to monitor costs post-adoption. This is unexpected given that the cost of using cloud services was so important to such a large proportion (86%) of the respondents' organisations in the decision-making process, and that cost saving was the most popular reason for using third-party cloud services.[31] In the case examples where models are used, the choice and number were influenced either by their customary use in the organisation, or by the knowledge of those taking the lead role in the decision-making process or in preparing the business case. Respondents made no explicit reference to using or being influenced by the literature available.

Comparing the most popular models used in practice that emerged from the study with the models reported in real or hypothetical scenarios in the literature, only Total Cost of Ownership featured in both research and practice. Although the small number of organisations or respondents using economic models limits any conclusions regarding their use, the data appears to indicate a preference for the simpler, perhaps more commonly used, comparisons of costs. In addition, perhaps the disparity between the large number of respondents indicating the importance of cost and the small number using cost models reflects ongoing assessments of cost but only in informal, unsystematic ways. This trend is concerning because scholarship examined in the literature (see above) developed complex models exactly because the simpler cost comparisons were found to be inadequate. However, in the two case examples where financial models were used, the interviewees, both recordkeeping professionals, were very aware of the shortcomings of the models used. In *Case 1* the organisation's focus on the financial aspect was balanced by their use of a non-financial model to ensure functional requirements for the archive collection were met. In *Case 3*, the interviewee confirmed the point made by Reichmanthat, [32] whilst Total Cost of Ownership (the sum of *all* costs – direct and indirect) is a good approach, it is difficult to use accurately in practice.

Lessons learned

The experiences of the case example organisations provide a lens on the issue of cloud decision-making not available in research identified in the literature. The issues they encountered offer a set of 'lessons learned' to inform other recordkeeping professionals considering cloud storage for records. These lessons cover cloud storage costing generally and the use of models in the process specifically.

First, the use of a costing model can be effective in supporting the business case but may not be adequate for longer-term predictions. *Case 1* was able to demonstrate that the economics of cloud storage made sense in the short to medium term, however, was open about this in the long term, since volume will only increase. *Case 5* also had concerns about the difficulty in predicting future storage and usage costs. *Case 3* showed that it is essential to quantify the *current* state and not a nominal future state because the risks of using the cloud may not be any different to the risks currently being faced. Ultimately, the models themselves may need further development.

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

Second, it is essential to identify all costs, not just technical ones, 'otherwise there may be no money to provide the records management service, only the platform' (*Case 3*). Hidden costs include the time people give 'freely' to ensure something gets done and 'behind-the-scenes people costs (including policies, maintaining security and access models, the ontological structure, the business classification model and its implementation, metadata, administration, training etc.) [which] are a significant proportion of the whole, and may amount to more than is saved on the technical costs' (*Case 3*). The cloud may be a more or less expensive option; for example hardware costs may be less but the cost of copies may be greater.[33] *Case 3* was still investigating the cost but thought it may be higher since the risks are greater when access is federated to others and consequently tight access and security controls are needed which has cost implications. These responses reveal cost considerations in a number of areas that are not captured by the existing models, suggesting areas for further research and development.

Third, cost modelling needs to be considered on a case-by-case basis, with a full understanding of the particular scenario. Context and usage are critical.[34] For example, in *Case 1* the expectation was that the organisation would be moving more, if not all, of its information storage into the cloud, thereby changing the economics of the in-house option. The scenario involved a significant volume of content but limited usage and only by recordkeeping professionals, making it relatively easy to predict; the economics would be very different if it were supporting public access and heavy usage demand, particularly of uploads and downloads. In other cases, cloud services can offer *'a viable means of entry' (Case 1)* for smaller organisations for whom the conventional route, with big up-front capital costs, would be unrealistic.

Fourth, these examples raise questions about the potential impact of moving from capital expenditure (CAPEX) to ongoing operating expenditure (OPEX) in using cloud storage services, something accounted for in Walker, Brisken and Romney's early model.[35] Research on digital preservation in the cloud suggests this is 'not simply a case of funding being reallocated from capital to operating – a much more complicated scenario is at play... it's ironic… there is this big push to consume things as a service which everyone knows is moving you down an op ex route. And the financial models aren't there to let you do it'.[36] There are models but they are not all well-known and/or adequate. The case examples here were at too early a stage to be able to assess the impact fully. For *Case 1*, it was difficult to assess partly because it was their only experience and there was no previous in-house system to benchmark against. For *Case 3*, if their IT Department's monitoring of storage cost were to show an unsustainable increase, individual business units could be approached to discuss reduction. IT's approach would likely involve indiscriminately cutting content to reduce cost, i.e. failing to consider the legal, regulatory and business requirements for records retention, demonstrating a disconnect between IT and records management.

This leads to the final lesson that, whilst recordkeeping professionals are often involved in the decision-making process (38 responses) they are more often not the leading voice in decisions about cloud storage (only 4 respondents played the lead role). This was borne out in the case examples. Given the functional requirements for records storage it is worrying that they are not automatically involved or take the lead. It is in records that one can find the full suite of

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

reasons why cloud storage would be considered in the first place – security, cost of storage, need for sharing, economies of scale, infrastructure availability and costs, expertise, trust etc. Articulating and understanding these reasons, including those that are complex, subjective, temporal, situation-specific and difficult to measure (e.g. functional requirements), should lead to more robust economic and related decision-making models that can be applied with a greater level of confidence in any information generating environment where cloud storage options are being considered. Recordkeeping and other information professionals have a key role and 'must be prepared to be assertive … based on their mandate to act as stewards of information as an authoritative resource'.[37]

Addressing the emergent issues

Two key issues emerged from the study: first that models are available for estimating the cost of storing records in the cloud but are not widely used, and second that the models are inadequate for estimating longer term records storage costs. Recordkeeping professionals have an important role in addressing both issues.

Archivists and records managers need to be aware of and understand the various costing models, and ensure the most appropriate ones are used in the cloud storage decision-making process. Checklists and guides for cloud usage aimed explicitly at recordkeeping professionals can help with awareness raising;[38] however, they do not appear to explicitly cover cost modelling. Therefore they could be further developed. For current professionals deeper knowledge and understanding of economic issues and models can be gained either through discussion with colleagues (e.g. contract managers, financial or business managers), or by exploring the literature reviewed earlier, or through other continuing professional development opportunities. Professional organisations could play a role here in offering such opportunities and educators could include economic issues and models in course curricula for new professionals.

With a greater awareness and understanding of costing models, together with their knowledge of a given context, scenario and requirements, recordkeeping professionals could work with others (e.g. statisticians, economics, business modellers) to develop more effective ones. Such models would consider all of the functional requirements, hidden (human) costs and risks. They could be tested in 'live' situations through collaborative research fostered by, for example, universities or professional organisations. Testing would need to consider other related factors such as issues of trust in the ongoing viability and sustainability of cloud services, which also emerged from our study.[39] This might feed back into research and lead to further development of costing models based on real data, together with recommended strategies for monitoring and re-modelling costs at appropriate intervals. Such work would help information professionals to position themselves more effectively, establish more effective partnerships, and become more actively involved in the decision-making process.

Conclusions

The rapid increase in the volume of digital data and information and the potential to 'keep everything', counter-balanced with queries about the validity of Kryder's Law and inevitable budget pressures, [40] alongside the environmental (green) agenda, mean the economics of cloud storage is an important issue for recordkeeping professionals. Models for assessing and

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

monitoring the economic viability and sustainability of cloud storage services should be seriously considered, since these issues are key to trusting in cloud storage services and in service providers to give a fair deal and continue to uphold their agreement.[41] Despite research into the value and development of models, there is little published evidence of their application in practice. The study discussed here helps to fill that gap. While focusing on records storage, which presents unique, complex challenges, the findings have wider relevance for data which are part of the data-information-records-archives continuum.

If cost is a primary driver toward use of the cloud, organisations should assess if the service is actually cost effective. Crucial to archival concerns, recordkeeping professionals must correctly determine the costs of cloud storage in order to trust that the service will be sustainable over time. For records, as a form of digital information with special characteristics and particular preservation needs, economic sustainability is one important part of a larger nexus of trust issues which recordkeeping professionals navigate in the cloud decision-making process. Records professionals have broader concerns related to the sustainability of the storage service, such as issues with contractual terms and vendor lock-in, and the ability of the service to meet records requirements, which influence the decision-making process alongside cost issues.[42]

Although the extent of data collected means that conclusions are limited, it does provide a range of experiences and lessons learned. Currently, costing/economic models are not widely used in the cloud storage service context, meaning current decision-making practices may not be adequate. Calculating costs is challenging and complicated by the number of variables to consider,[43] accuracy of future growth forecasts,[44] hidden costs and contractual commitments, which may be difficult to keep if funding levels, mechanisms and/or policies change. Work is needed to turn scholars' models into tools that practitioners can use in various information storage contexts.

Multiple players (recordkeeping and IT professionals; business managers; cloud service providers; modellers etc.) should be involved in the application of relevant models. For recordkeeping professionals it is of particular concern since cloud platform and software services result in records creation and storage ‘by default’. If they do not proactively initiate the modelling of costs, there is a danger that digital records of all kinds will be vulnerable to ad hoc or inappropriate retention management. Yet this study highlights a gap between recordkeeping professionals and the other groups that is important to close.

In the paper world, recordkeeping professionals are very familiar with the cost implications of records storage facilities and capacities; they need to translate this expertise into the digital world for records (and data) stored in cloud-based business systems. Together with their expertise in the functional requirements for medium to longer-term storage of records, they could work with other stakeholders to undertake research developing more relevant, reliable, complete and effective models. Once such models are available they can be adopted more widely by recordkeeping and other information professionals and by those educating new professionals. In the interim, archivists and records managers should at least take the initiative to discuss cost modelling with their IT colleagues, who more often appear to take the lead, bring their understanding to bear and move more ‘centre stage’.

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

References

[1] IDC, 'Growth in Cloud IT Infrastructure Spending Will Accelerate in 2017 Driven by Public Cloud Datacenters and On-Premises Private Cloud Environments', Press release 13 Jan, 2017, available at <https://www.idc.com/getdoc.jsp?containerId=prUS42056416> accessed 14 Nov 2017.
[2] Forrester Research Inc., 'File Storage Costs Less in the Cloud than In-House', 2011, available at <http://www.forrester.com/report/File+Storage+Costs+Less+In+The+Cloud+Than+InHouse/-/E-RES57696>, accessed 1 March 2017; Gartner Inc., 'Case Studies in Cloud Computing', 2011, available at <http://www.gartner.com/doc/1761616/case-studies-cloud-computing>, accessed 2 May 2017.
[3] K Stuart and D Bromage, 'Current State of Play: Records Management and the Cloud', *Records Management Journal*, vol. 20, no. 2, 2010, pp. 217–225; A Brown and C Fryer, 'Achieving Sustainable Digital Preservation in the Cloud', 2nd Annual Conference of the International Council on Archives, Girona, Spain, 11–15 Oct, 2014, available at <http://www.girona.cat/web/ica2014/ponents/textos/id87.pdf>; G Oliver, 'Digital Preservation in the Cloud (AA01), InterPARES Trust project', 2nd Annual Conference of the International Council on Archives, Girona, Spain, 11–15 Oct, 2014. G Oliver and S Knight, 'Storage is a Strategic Issue: Digital Preservation in the Cloud', *D-Lib Magazine*, vol. 21, no. 3/4, 2015, available at <http://www.dlib.org/dlib/march15/oliver/03oliver.html>, accessed 2 May 2017; O Zander, 'Preserving 40 terabytes per day. On-premises, cloud … or both?' 2nd Annual Conference of the International Council on Archives, Girona, Spain, 11–15 Oct, 2014.
[4] ISO, 'ISO 15489-1 Information and Documentation – Records Management. Part 1: Concepts and Principles', ISO, Geneva, 2016.
[5] J McLeod and B Gormly, 'Economic Models for Storage of Records in the Cloud (StaaS) – A Critical Review of the Literature (EU18), InterPARES Trust project', 2015, available at <*https://interparestrust.org/assets/public/dissemination/EU18_20150713_CloudEconomicsLitReview_FinalReport.pdf*>, accessed 2 May 2017. See appendix (pp. 13–14) for full details of the databases and websites through which literature review sources were found.
[6] E Walker, 'The Real Cost of a CPU Hour', *Computer*, vol. 42, no. 4, 2009, pp. 35–41.
[7] E Walker, W Brisken and J Romney, 'To Lease or not to Lease from Storage Clouds. *Computer*, vol. 43, no. 4, 2010, pp. 44–50
[8] Walker, Brisken and Romney, p. 49.
[9] L Mastroeni and M Naldi, 'Storage Buy-or-Lease Decisions in Cloud Computing under Price Uncertainty', Paper presented at 7th EURO-NGI Conference on Next Generation Internet (NGI), Kaiserslautern, 27–29 Jun 2011; L Mastroeni and M Naldi, 'Long-range Evaluation of Risk in the Migration to Cloud Storage', in 13th IEEE Conference on Commerce and Enterprise Computing, CEC 2011, pp. 260–266, IEEE, Luxembourg; L Mastroeni and M Naldi, 'Analysis of Cloud Storage Prices', *arXiv preprint series, CoRR*, 2012, preprint no. 1207.6011v1; M Naldi and L Mastroeni, 'Cloud Storage Pricing: A Comparison of Current Practices', Paper presented at 2013 International Workshop on Hot topics in Cloud Services, 4th ACM/SPEC International Conference on Performance Engineering, Prague, 20–24 Apr 2013 doi: 10.1145/2462307.2462315; M Naldi and L Mastroeni, 'Economic Decision Criteria for the Migration to Cloud Storage', *European Journal of Information Systems*, vol. 25, no. 1, 2016, pp. 16–28. doi: 10.1057/ejis.2014.34. First available online 30/9/2014.
[10] O Mazhelis, G Fazekas and P Tyrvainen, 'Impact of Storage Acquisition Intervals on the Cost-Efficiency of the Private vs. Public Storage', Paper presented at 5th IEEE International Conference on Cloud Computing (CLOUD), Honolulu, HI, 24–29 Jun, 2012 doi: 10.1109/CLOUD.2012.101; G Laatikainen, O Mazhelis, and P Tyrväinen, 'Role of Acquisition Intervals in Private and Public Cloud Storage Costs', *Decision Support Systems*, vol. 57, 2014, pp. 320–330 doi: 10.1016/j.dss.2013.09.020
[11] A Khajeh-Hosseini, D Greenwood, JW Smith and I Sommerville, 'The Cloud Adoption Toolkit: Supporting Cloud Adoption Decisions in the Enterprise', *Software: Practice and Experience*, vol. 42, no. 4, 2012, pp. 447–465 doi: 10.1002/spe.1072; J Wang, R Hua, Y Zhu, C Xie, P Wang and W Gong, 'C-

IRR: An Adaptive Engine for Cloud Storage Provisioning Determined by Economic Models with Workload Burstiness Consideration', Paper presented at 7th IEEE International Conference on Networking, Architecture and Storage (NAS), Xiamen, Fujian, 28–30 Jun, 2012 doi: 10.1109/NAS.2012.13; AK Dutta and R Hasan, 'How Much Does Storage Really Cost? Towards a Full Cost Accounting Model for Data Storage', in *Economics of Grids, Clouds, Systems, and Services, 10th International Conference: GECON 2013, Zaragoza, Spain, 18–20 Sep, 2013*, Springer International Publishing, pp. 29–43. doi: 10.1007/978-3-319-02414-1_3; A Reichman, 'File Storage Costs Less in the Cloud than In-House', 2011, available at <http://media.amazonwebservices.com/Forrester_File_Storage_Costs_Less_In_The_Cloud.pdf>, accessed 2 May 2017.

[12] DSH Rosenthal and DL Vargas, 'LOCKSS Boxes in the Cloud', 2012, available at <http://www.lockss.org/locksswp/wp-content/uploads/2012/09/LC-final-2012.pdf>, accessed 2 May 2017; DC Rosenthal, DSH Rosenthal, EL Miller, IF Adams, MW Storer and E Zadok, 'Toward an Economic Model of Long-Term Storage', 2012, available at <*http://static.usenix.org/events/fast/poster_descriptions/Rosenthaldescription.pdf*>, accessed 2 May 2017; DSH Rosenthal, DC Rosenthal, EL Miller, IF Adams, MW Storer and E Zadok, 'The Economics of Long-Term Digital Storage', 2012, available at <http://www.lockss.org/locksswp/wp-content/uploads/2012/09/unesco2012.pdf>; 4C (Collaboration to Clarify the Costs of Curation) (2013–15), '4C: Collaboration to Clarify the Costs of Curation', 2015, available at <http://www.4cproject.eu/>, accessed 2 May 2–17.

[13] McLeod and Gormly, 2015, p. 5-8.

[14] Walker, Brisken and Romney.

[15] C Walter, 'Kryder's Law', *Scientific American*, no. 293 (August), 2005, pp. 32–33 and Wikipedia entry https://en.wikipedia.org/wiki/Mark_Kryder#Kryder.27s_Law, accessed 2 May 2017.

[16] Rosenthal and Vargas, p. 4.; Rosenthal et al.

[17] Reichman, p.13.

[18] Laatikainen, Mazhelis and Tyrväinen; Mazhelis, Fazekas and Tyrväinen.

[19] Khajeh-Hosseini et al., p. 461–63 ; Dutta and Hasan, p. 39–40

[20] Mazhelis, Fazekas and Tyrväinen, p. 4; Laatikainen, Mazhelis and Tyrväinen, p.328.

[21] Walker, Brisken and Romney; Naldi and Mastroeni; Mastroeni and Naldi, *'Storage Buy-or-Lease Decisions in Cloud Computing under Price Uncertainty'*, Paper presented at 7th EURO-NGI Conference on Next Generation Internet (NGI), Kaiserslautern, 27—29 Jun 2011; DS Rosenthal et al.; Reichman.

[22] Wang et al.; Rosenthal and Vargas; Naldi and Mastroeni, 2013.

[23] Rosenthal and Vargas, p.7.

[24] RW Johnson and WG Lewellen, 'Analysis of Lease-or-Buy Decision', *Journal of Finance*, vol. 27, no. 4, 1972, pp. 815—823.

[25] Dutta and Hasan; DSH Rosenthal et al.

[26] J McLeod and B Gormly, 'Economic models for cloud storage decision-making: An investigation into the use of economic models for making decisions about using the cloud for records storage (EU20). InterPARES Trust Project', 2016, available at <*https://interparestrust.org/assets/public/dissemination/EU20_20160609_CloudEconomicModels_EUWorkshop8_FinalReport.pdf*>, accessed 2 May 2017.

[27] Available at <http://fluidsurveys.com/, accessed 2 March 2016.

[28] McLeod and Gormly, 2016.

[29] J McLeod and B Gormly, 'Using the cloud for records storage: Issues of trust'. *Archival Science*, vol. 17, no. 4, 2017, pp. 349-370. doi: 10.1007/s10502-017-9280-5

[30] McLeod and Gormly, 2016, p. 62 and 70.

[31] Ibid.

[32] Reichman, p. 13.

Archives and Manuscripts, 2018, 46:2, 174-192 DOI:10.1080/01576895.2017.1409125

[33] Union Solutions, 'Stop Comparing the Cost of Storage, Start Understanding Data Storage Services', 2016, available at <http://blog.unionsolutions.co.uk/understand-data-storage-services-cost>, accessed 2 May 2017.
[34] Ibid.
[35] Walker, Brisken and Romney.
[36] Oliver and Knight, p. 7.
[37] Ibid, p. 9.
[38] See for example: J Bushey, M Demoulin, E How and R McLelland, 'Checklist for Cloud Service Contracts', 2016, available at <https://interparestrust.org/assets/public/dissemination/NA14_20160226_CloudServiceProviderContracts_Checklist_Final.pdf>, accessed on 2 May 2017; National Archives of Australia, 'Cloud Computing and Information Management', no date, available at <http://www.naa.gov.au/records-management/agency/secure-and-store/cloud-computing/index.aspx#section3>, accessed 2 May 2017; New Zealand Government, 'Records Management and the Cloud. Version 1', 2015, available at <http://records.archives.govt.nz/assets/Archives-ResourcesandGuides-Quick-Guides/QUICK-GUIDE-Records-Management-and-the-Cloud-June-2015-final.docx>, accessed 2 May 2017; State Records of South Australia, 'Cloud Computing and Records Management. Guideline Version 1', 2015, available at <https://government.archives.sa.gov.au/sites/default/files/20150706%20Cloud%20Computing%20and%20Records%20Management%20Final%20V1.pdf>, accessed 2 May 2017.
[39]McLeod and Gormly, 2017
[40] DSH Rosenthal et al.
[41] McLeod and Gormly, 2017
[42] Ibid, for further discussion
[43] Khajech-Hosseini et al., p. 448.
[44] Laatikainen, Mazhelis and Tyrväinen, p. 321 and 327.